

09. 7. 2004

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年 7月11日
Date of Application:

出願番号 特願2003-273590
Application Number:
[ST. 10/C]: [JP 2003-273590]

REC'D: 26 AUG 2004

WIPO

PCT

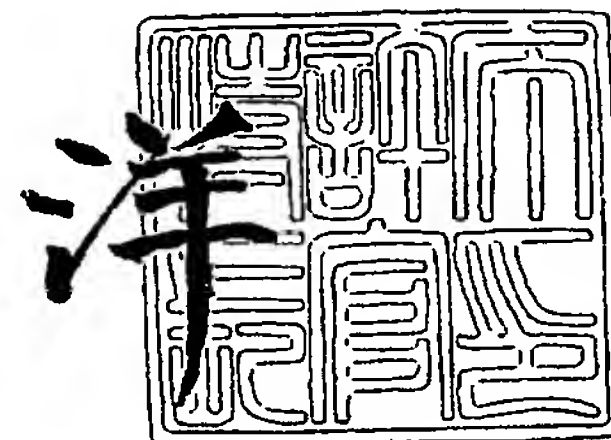
出願人 トヨタ自動車株式会社
Applicant(s):

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2004年 8月13日

特許庁長官
Commissioner,
Japan Patent Office

小川



BEST AVAILABLE COPY

出証番号 出証特2004-3072350

【書類名】 特許願
【整理番号】 2003-3374Z
【提出日】 平成15年 7月11日
【あて先】 特許庁長官殿
【国際特許分類】 G06T 1/00
【発明者】
 【住所又は居所】 愛知県豊田市トヨタ町1番地 トヨタ自動車株式会社内
 【氏名】 船山 竜士
【発明者】
 【住所又は居所】 愛知県豊田市トヨタ町1番地 トヨタ自動車株式会社内
 【氏名】 鄭 文皓
【発明者】
 【住所又は居所】 愛知県豊田市トヨタ町1番地 トヨタ自動車株式会社内
 【氏名】 大杉 雅道
【特許出願人】
 【識別番号】 000003207
 【氏名又は名称】 トヨタ自動車株式会社
【代理人】
 【識別番号】 100088155
 【弁理士】
 【氏名又は名称】 長谷川 芳樹
【選任した代理人】
 【識別番号】 100089978
 【弁理士】
 【氏名又は名称】 塩田 辰也
【手数料の表示】
 【予納台帳番号】 014708
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

単眼カメラで顔を撮像し、撮像した動画像から視線を求める画像処理装置であって、
顔画像上の瞳中心の 3 次元構造を静的パラメータと動的パラメータで定義し、前記静的
パラメータ及び前記動的パラメータを推定することによって視線を求めることを特徴とす
る画像処理装置。

【請求項 2】

前記静的パラメータは、カメラ座標系における瞳の奥行きであることを特徴とする請求項
1 に記載する画像処理装置。

【請求項 3】

前記動的パラメータは、眼球の回転パラメータであることを特徴とする請求項 1 又は請求
項 2 に記載する画像処理装置。

【請求項 4】

前記眼球の回転パラメータは、眼球の座標系での 2 つの座標軸に対して回転可能である 2
自由度であることを特徴とする請求項 3 に記載する画像処理装置。

【書類名】 明細書

【発明の名称】 画像処理装置

【技術分野】

【0001】

本発明は、単眼カメラで撮像した顔の動画像に画像処理を施す画像処理装置に関するものであり、特に、顔の動画像から視線を推定する画像処理装置に関する。

【背景技術】

【0002】

画像処理には、カメラで顔を撮像し、その動画像から顔自体の姿勢や顔の各部の動き等の検出を行うものがある。このような画像処理には、眼の視線を求めるものがある（非特許文献1参照）。非特許文献1には、単眼カメラで顔を撮像し、その動画像から顔を追跡しながら視線を求めていることが記載されている。その視線を求める方法としては、顔の多数の特徴点をカルマンフィルタリングで求め、その求めた特徴点を用いて顔の姿勢及び視線を求めている。

【非特許文献1】 J. Heinzmann and A. Zelinsky, "3-D facial pose and gaze point estimation using a robust real-time tracking paradigm", IEEE Int. Workshop on Automatic Face and Gesture Recognition, pp142-147, 1998

【発明の開示】

【発明が解決しようとする課題】

【0003】

上記した視線を求める方法では、顔姿勢や視線のパラメータを直接フィルタリングするのではなく、特徴点にのみフィルタリングを行っているので、視線に対しての最適な推定を行えない。また、この方法では、眼の両端の特徴点と瞳の中心点との相対的な位置変化を用いて視線を求めているので、特定の特徴点（すなわち、眼の両端の特徴点）に依存し、視線を求めるのには限界がある。さらに、眼の両端の特徴点と瞳の中心点の相対的な位置変化を用いて視線を求めるが、カメラモデルのアフィン変換[Affine Transformation]を前提としている。そのため、この方法は、一般的な透視投影変換[Perspective Transformation]の下では成り立たないので、応用範囲が狭い。

【0004】

そこで、本発明は、視線を高精度に求めることができる画像処理装置を提供することを課題とする。

【課題を解決するための手段】

【0005】

本発明に係る画像処理装置は、単眼カメラで顔を撮像し、撮像した動画像から視線を求める画像処理装置であって、顔画像上の瞳中心の3次元構造を静的パラメータと動的パラメータで定義し、静的パラメータ及び動的パラメータを推定することによって視線を求めることを特徴とする。

【0006】

この画像処理装置では、単眼カメラで顔を撮像した動画像に対して、顔画像上の瞳中心点における3次元構造を静的パラメータと動的パラメータによって定義する。さらに、画像処理装置では、フィルタリング等によって静的パラメータ及び動的パラメータを推定し、この推定値から視線方向を求める。このように、画像処理装置では、実際に起こりうる瞳の動きを静的パラメータや動的パラメータを用いてモデリングしているので、高精度に視線を推定することができる。また、この画像処理装置は、単眼カメラで、無制限の顔の動きに対応しながら視線の最適推定を行うことができる。

【0007】

なお、動画像は、時間的に連像した画像である。静的パラメータは、決まった値を元々

有しているパラメータであり、例えば、眼球や瞳に関する値である。動的パラメータは、決まった値を有していないパラメータであり、例えば、顔や眼球の動きで変わる値である。

【0008】

本発明の上記画像処理装置では、静的パラメータは、カメラ座標系における瞳の奥行きであると好適である。さらに、本発明の上記画像処理装置では、動的パラメータは、眼球の回転パラメータであると好適であり、特に、眼球の回転パラメータは、眼球の座標系での2つの座標軸に対して回転可能である2自由度であると好適である。

【発明の効果】

【0009】

本発明によれば、高精度に視線を求めることができる。

【発明を実施するための最良の形態】

【0010】

以下、図面を参照して、本発明に係る画像処理装置の実施の形態を説明する。

【0011】

本実施の形態では、本発明に係る画像処理装置を、単眼カメラで撮像した顔の動画像に画像処理を施す画像処理システムの画像処理装置に適用する。本実施の形態に係る画像処理装置は、コンピュータ上に構成され、少なくとも顔の動画像から視線を推定することができる装置である。なお、本実施の形態では、顔の動きには制限がなく、右眼の視線方向と左眼の視線方向とは同じ方向として視線の推定を行う。

【0012】

図1を参照して、本実施の形態である画像処理システム1について説明する。図1は、本実施の形態に係る画像処理システムの構成図である。

【0013】

画像処理システム1は、透視投影変換のカメラモデルを用い、非接触で視線推定を行う。画像処理システム1では、画像取得装置2が画像処理装置3に接続され、画像取得装置2によって顔を撮像し、顔の動画像データを取得する。そして、画像処理システム1では、画像処理装置3においてその取得した動画像データから視線の最適推定を行う。画像処理システム1では、必要に応じて、画像処理の過程や結果を出力するためにモニタ（図示せず）等の出力手段が画像処理装置3に接続される。なお、画像処理システム1では、少なくとも視線の推定を行うが、必要に応じて、顔の姿勢の推定等の顔に対する他の画像処理も行う。

【0014】

画像取得装置2は、CCD[Charge Coupled Device]カメラ（図示せず）及びデータ変換装置（図示せず）からなる。CCDカメラは、単眼カメラであり、撮像素子としてCCDを用いている。CCDカメラでは、カメラのレンズから入ってきた光をCCDに集め、CCDにおいて光の情報をデジタルデータに変換して画像データとして出力する。また、CCDカメラでは、撮像対象を時間的に連続して撮像し、時間的に連続した画像（動画像）データを出力する。データ変換装置は、CCDカメラから入力された動画像データの形式を画像処理装置3で取り扱うことができる形式に変換する。画像取得装置2では、CCDカメラで被験者の顔全体を撮像し、データ変換装置でその動画像データを所定の形式に変換して画像処理装置3に出力する。

【0015】

図1～図6を参照して、画像処理装置3について説明する。図2は、カメラ座標系である。図3は、顔が固定の場合のカメラ座標系、顔座標系、眼球座標系、視線座標系である。図4は、眼球モデルである。図5は、Dynamic Converging-Connected Gaze Modelである。図6は、顔が動いた場合のカメラ座標系、顔座標系、眼球座標系、視線座標系である。

【0016】

画像処理装置3は、パーソナルコンピュータ等のコンピュータに構成され、3次元モデ

ル初期化部 10 及び推定部 11 を備えている。3次元モデル初期化部 10 及び推定部 11 は、画像処理装置 3 用のアプリケーションプログラムをコンピュータ上で実行することによって構築される。

【0017】

なお、画像処理装置 3 では、 $N+2$ 個 ($i=1, 2, \dots, N+2$) の特徴点を取り扱うが、そのうちの N 個 ($i=1, 2, \dots, N$) が瞳の中心点以外の顔の特徴点であり、2 個 ($i=N+1, N+2$) が左右の瞳の中心点である。また、画像処理装置 3 では、図 3 及び図 6 に示す 4 つの座標系を定義しており、 (X_c, Y_c, Z_c) が画像処理を行う上での基準の座標系であるカメラの座標系であり、 (X_h, Y_h, Z_h) が顔の座標系であり、 (X_e, Y_e, Z_e) が眼球の座標系であり、 (X_g, Y_g, Z_g) が視線 (瞳) の座標系である。

【0018】

3次元モデル初期化部 10 について説明する。3次元モデル初期化部 10 では、初期フレームに基づいて、顔の特徴点と瞳の特徴点 (中心点) の 3次元構造を定義する。初期フレームは、動画像における時間的に最初のフレームであり、画像処理を行う上での基準となるフレームである。初期フレームでは、顔が固定であり、顔座標系 (X_h, Y_h, Z_h) がカメラ座標系 (X_c, Y_c, Z_c) に一致している。

【0019】

3次元モデル初期化部 10 では、まず、初期フレームの画像から N 個の顔の特徴点 (x_i, y_i, z_i) ($i=1, 2, \dots, N$) を抽出する。そして、3次元モデル初期化部 10 では、抽出した顔の特徴点 (x_i, y_i, z_i) から式 (1) に示す顔の 3次元構造を定義する。顔の特徴点 (x_i, y_i, z_i) は、カメラ座標上の位置である。

【0020】

【数 1】

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = \begin{pmatrix} u_i \\ v_i \\ 0 \end{pmatrix} + \alpha_i \begin{pmatrix} u_i \beta \\ v_i \beta \\ 1 \end{pmatrix} \quad \dots (1)$$

$i=1, 2, \dots, N$

x_i, y_i, z_i : カメラ座標系上の顔の特徴点の座標 (位置)

u_i, v_i : 顔の特徴点の画像面 (CCD) 上における座標 (観測値)

α_i : 画像面 (CCD) から特徴点までの距離 (特徴点の奥行き)

$\beta = f^{-1}$, f : カメラの焦点距離

式 (1) において、 $(u_i, v_i, 0)$ は、顔の特徴点を CCD 上に投影した観測点であり、初期フレームで求めた後はその値に固定される。 α_i は、顔の特徴点 (x_i, y_i, z_i) から CCD までの距離であり、特徴点の奥行きを示している。 β は、CCD カメラの焦点 (f, f^{-1}) である。

【0021】

なお、図 2 には、カメラ座標系を示しており、 $(0, 0, -f)$ が CCD カメラの焦点 (カメラのレンズの中心) であり、焦点距離 f 離れたところに画像面 (すなわち、レンズからの光を受ける CCD) が位置している。カメラ座標系の原点 $(0, 0, 0)$ は、画像面上にある。画像面には、顔の特徴点 (x_c, y_c, z_c) が投影されると、観測点 $(u, v, 0)$ となる。カメラ座標系の Z 軸方向において、原点 $(0, 0, 0)$ から特徴点 (x_c, y_c, z_c) までの奥行きが α となる。カメラ座標系では、焦点距離 f と奥行き α とに分離されている。

【0022】

次に、3次元モデル初期化部 10 では、初期フレームの画像から左右の瞳の中心点 (u_i, v_i) ($i=N+1, N+2$) を抽出する。そして、3次元モデル初期化部 10 では、抽出した瞳の中心点 (u_i, v_i) 、眼球の回転角、カメラ座標系上の奥行きで瞳の中心点

の3次元構造を定義する。そのために、3次元モデル初期化部10では、カメラ座標系 (X_c, Y_c, Z_c)、顔座標系 (X_h, Y_h, Z_h)、眼球座標系 (X_e, Y_e, Z_e)、視線座標系 (X_g, Y_g, Z_g) を定義する (図3参照)。なお、画像処理装置3では、初期フレームにおいて、眼球座標系 (X_e, Y_e, Z_e) とカメラ座標系 (X_c, Y_c, Z_c) 及び顔座標系 (X_h, Y_h, Z_h) とはZ軸が平行で向きが逆方向と仮定している。

【0023】

顔が固定で眼球が回転した場合、視線座標系 (X_g, Y_g, Z_g) は、顔座標系 (X_h, Y_h, Z_h) に対して式 (2) のように表される。顔座標系 (X_h, Y_h, Z_h) (カメラ座標系 (X_c, Y_c, Z_c)) から視線座標系 (X_g, Y_g, Z_g) への変換 T_h^g は、顔座標系 (X_h, Y_h, Z_h) (カメラ座標系 (X_c, Y_c, Z_c)) から眼球座標系 (X_e, Y_e, Z_e) への変換 T_h^e と眼球座標系 (X_e, Y_e, Z_e) から視線座標系 (X_g, Y_g, Z_g) への変換 T_e^g との積によって表される (図3参照)。

【0024】

【数2】

$$T_h^g = T_h^e T_e^g = \begin{pmatrix} R_{Xh, \pi} & t_h \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R_e(\phi, \theta) & 0 \\ 0 & 1 \end{pmatrix} \dots (2)$$

$$t_h = \begin{pmatrix} u_i(1 + \beta \alpha_i) \\ v_i(1 + \beta \alpha_i) \\ \alpha_i + r_{eye} \end{pmatrix} \dots (2-1)$$

$$R_{Xh, \pi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \pi & -\sin \pi \\ 0 & \sin \pi & \cos \pi \end{pmatrix} \dots (2-2)$$

$$R_e(\phi, \theta) = \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ -\sin \phi \sin \theta & \cos \phi & \sin \phi \cos \theta \\ -\cos \phi \sin \theta & -\sin \phi & \cos \phi \cos \theta \end{pmatrix} \dots (2-3)$$

$i = N+1, N+2$ (左右の瞳の中心点)

T_h^g : 視線座標系の顔座標系に対する変換 (4×4の行列)

T_h^e : 眼球座標系の顔座標系に対する変換 (4×4の行列)

T_e^g : 視線座標系の眼球座標系に対する変換 (4×4の行列)

t_h : 眼球座標系の顔座標系に対する並進移動 (3×1の並進行列)

$R_{Xh, \pi}$: 眼球座標系の顔座標系に対する回転 (3×3の回転行列)

$R_e(\phi, \theta)$: 視線座標系の眼球座標系に対する回転 (3×3の回転行列)

r_{eye} : 眼球的半径

ϕ, θ : 眼球座標系における瞳の中心点の動き (眼球的回転) を示す回転角度

u_i, v_i : 瞳の中心点の画像面 (CCD) 上における座標 (観測値)

α_i : 画像面 (CCD) から瞳の中心点までの距離 (瞳の奥行き)

変換 T_h^e は、式 (2-1) による並進移動 t_h と式 (2-2) による回転 $R_{Xh, \pi}$ によって表される (図3参照)。並進移動 t_h は、顔座標系 (X_h, Y_h, Z_h) (カメラ座標系 (X_c, Y_c, Z_c)) から眼球座標系 (X_e, Y_e, Z_e) への原点の移動である。回転 $R_{Xh, \pi}$ は、Z軸に対する180°回転である。変換 T_e^g は、式 (2-3) による回転 R_e によって表される (図3参照)。回転 R_e は、図4に示すように、眼球的回転による瞳の中心点の I_0 から I への回転である。この瞳の中心点である I は、式 (3) で表される。

【0025】

【数 3】

$$\begin{pmatrix} I \\ 1 \end{pmatrix} = T_h^g \begin{pmatrix} I_0 \\ 1 \end{pmatrix} = T_h^g \begin{pmatrix} 0 \\ 0 \\ r_{eye} \\ 1 \end{pmatrix} \dots (3)$$

I : 顔座標系からの瞳の中心点

I_0 : 初期フレームでの眼球座標系における瞳の中心点

I_0 は、初期フレームにおける瞳の中心点である。 I_0 は、図4に示すように、眼球座標系 (X_e, Y_e, Z_e) においてZ軸上にあり、Z軸上における眼球の半径 r_{eye} に位置する。眼球の回転による瞳の中心点の I_0 から I への移動は、図4に示すように、2つの回転角 θ, ϕ によって規定される。つまり、眼球座標系 (X_e, Y_e, Z_e) のうちの Y_e 軸と X_e 軸に対して回転する角度によって規定されている。この2つの軸に対して回転可能となっている回転角を動的パラメータとする。 θ, ϕ は、 Y_e 軸と X_e 軸に対して回転可能であることから2自由度である。

【0026】

図4に示す I_0 を中心とした r_{iris} は、瞳の半径である。瞳の半径 r_{iris} を画像面 (CD) 上に投影すると、式 (4-1) に示すようになる。さらに、眼球の半径 r_{eye} を画像面上に投影した瞳の半径 s_i によって表すと、式 (4-2) に示すようになる。

【0027】

【数 4】

$$s_i = \frac{r_{iris}}{1 + \beta \alpha_i} \dots (4-1)$$

$$r_{eye} = \frac{r_{eye}}{r_{iris}} \times \frac{r_{iris}}{s_i} \times s_i = \lambda s_i (1 + \beta \alpha_i) \dots (4-2)$$

$i = N+1, N+2$

s_i : 画像面 (CCD) 上に投影した瞳の半径

r_{iris} : 瞳の半径

λ : 眼球の半径と瞳の半径との半径比

式 (4-2) における λ は、眼球と瞳との半径比であり、一般的に人によって差が少ないので一定値と仮定することができる。さらに、左右の眼球の半径も同じ値と仮定すると、式 (5) が成立する。

【0028】

【数 5】

$$s_{N+1} (1 + \beta \alpha_{N+1}) = s_{N+2} (1 + \beta \alpha_{N+2}) \dots (5)$$

式 (5) において、下付き文字の $N+1, N+2$ のうちの一方が左の眼球 (瞳) を示し、他方が右の眼球 (瞳) を示す。

【0029】

3次元モデル初期化部10では、式 (2) ~ 式 (5) から、式 (6) に示す顔が固定の場合の眼球運動による瞳の中心点の3次元構造を定義する。

【0030】

【数 6】

$$I = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = \begin{pmatrix} (u_i + \lambda s_i \sin \theta) (1 + \beta \alpha_i) \\ (v_i + \lambda s_i \sin \phi \cos \theta) (1 + \beta \alpha_i) \\ \alpha_i + \lambda s_i (1 - \cos \phi \cos \theta) (1 + \beta \alpha_i) \end{pmatrix} \dots (6)$$

$i = N+1, N+2$

$$\alpha_{N+2} = \frac{1}{\beta} \left(\frac{s_{N+1}}{s_{N+2}} - 1 \right) + \frac{s_{N+1}}{s_{N+2}} \alpha_{N+1} \dots (6-1)$$

x_i, y_i, z_i : 顔が固定の場合のカメラ座標系上の瞳の中心点の座標 (位置)

式 (6) において、瞳中心点の観測値 (u_i, v_i) は、初期フレームで求められた固定値である。半径比 λ は、実際には人によって異なるがその差が少ないので、人体工学によって求められる一定値を用いる。ちなみに、半径比 λ は、実際に測定した値を用いてもよい。回転角 (θ, ϕ) は、眼球の回転運動によって変わり、未知である。回転角 (θ, ϕ) は、値が決まっていないので、動的パラメータとする。焦点距離 β は、CCDカメラによって変わり、未知である。瞳中心点の奥行き α_i は、決まった値を元々有しているが、未知である。奥行き α_i は、ある決まった値に収束するので、静的パラメータとする。

【0031】

なお、瞳中心点の奥行き α_i は、式 (5) から、式 (6-1) に示すように、左右の瞳の一方の奥行き α_{N+2} を他方の奥行き α_{N+1} で表すことができる。したがって、他方の奥行き α_{N+1} を推定すれば、一方の奥行き α_{N+2} を求めることができる。

【0032】

推定部 11 について説明する。推定部 11 では、顔の運動モデルと眼球の運動モデルを定義し、この 2 つの運動モデルから Dynamic Converging-Connected Gaze Model (以下、DCCGM と記載する) を生成する。そして、推定部 11 では、DCCGM に対して拡張カルマンフィルタリングし、視線を推定する。

【0033】

顔の動きをカメラ座標系 (X_c, Y_c, Z_c) に対して回転と並進移動で表すと、顔の特徴点 ($x_{ci}, y_{ci}, \beta z_{ci}$) は、式 (7-1) となる。

【0034】

【数 7】

$$\begin{pmatrix} x_{ci} \\ y_{ci} \\ \beta_{ci} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \beta \end{pmatrix} \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} + R_c \bar{m}_i \quad \dots (7-1)$$

$$\bar{m}_i = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = \begin{pmatrix} u_i + \alpha_i u_i \beta \\ v_i + \alpha_i v_i \beta \\ \beta \end{pmatrix} \quad \dots (7-1-1)$$

$$\bar{o}_i = \begin{pmatrix} \frac{x_{ci}}{1 + \beta z_{ci}} \\ \frac{y_{ci}}{1 + \beta z_{ci}} \end{pmatrix} \quad \dots (7-2)$$

$$\bar{O} = \begin{pmatrix} \bar{o}_1 \\ \bar{o}_2 \\ \vdots \\ \bar{o}_N \end{pmatrix} \quad \dots (7-3)$$

$i = 1, 2, \dots, N$

x_{ci}, y_{ci}, z_{ci} : 顔が動いた場合の顔の特徴点をカメラ座標系上に投影した座標

t_x, t_y, t_z : 顔座標系のカメラ座標系に対する並進移動量

R_c : 顔座標系のカメラ座標系に対する回転 (3×3 の回転行列)

\bar{o}_i : (x_{ci}, y_{ci}, z_{ci}) を画像面 (CCD) 上に投影した観測ベクトル

\bar{O} : 全ての顔の特徴点を画像面 (CCD) 上に投影した観測ベクトル

式 (7-1) において、(t_x, t_y, t_z) はカメラ座標系 (X_c, Y_c, Z_c) からの 3×1 の並進行列であり、 R_c はカメラ座標系 (X_c, Y_c, Z_c) からの 3×3 の回転行列であり、 m_i ベクトルは式 (1) から式 (7-1-1) で表すことができる。

【0035】

この顔の特徴点 ($x_{ci}, y_{ci}, \beta z_{ci}$) を画像面 (CCD) 上に投影した \bar{o}_i ベクトル (観測ベクトル) は、式 (7-1) から式 (7-2) で表すことができる。ここで、全ての顔の特徴点 ($i = 1, 2, \dots, N$) に対する観測ベクトルである \bar{O} ベクトルは、式 (7-3) で表すことができる。このように、推定部 11 では、顔の特徴点に対する観測モデルを表す。

【0036】

そして、推定部 11 では、この顔の運動モデルの状態変数である p ベクトルを式 (8-1) で定義する。 p ベクトルは、顔の姿勢と構造パラメータで構成される。さらに、推定部 11 では、顔の運動モデルの状態空間表現を式 (8-2) で定義する。

【0037】

【数 8】

$$\bar{p} = \begin{pmatrix} t_x \\ t_y \\ t_z \\ w_x \\ w_y \\ w_z \\ \beta \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{N+1} \end{pmatrix} \quad \dots (8-1)$$

$$\bar{p}_t = A^{(1)} \bar{p}_{t-1} + \mu_t^{(1)} \quad \dots (8-2)$$

\bar{p} : 顔の運動モデルの状態変数

w_x, w_y, w_z : 前フレームからの各軸の回転量 (回転行列 R_c に関する量)

\bar{p}_t : 現時点での顔の運動モデルの状態変数

\bar{p}_{t-1} : 一時刻前での顔の運動モデルの状態変数

$A^{(1)}$: 状態遷移行列

$\mu_t^{(1)}$: ノイズ (ゼロの平均と $Q^{(1)}$ で構成されるガウシアン分布のランダムベクトル)

$Q^{(1)}$: ノイズの分散共分散行列

式 (8-1) において、 (t_x, t_y, t_z) は顔座標系 (X_h, Y_h, Z_h) のカメラ座標系 (X_c, Y_c, Z_c) に対する並進移動量であり、 (w_x, w_y, w_z) は前フレームからのカメラ座標系 (X_c, Y_c, Z_c) に対する各軸の回転量であり、 β は CCD カメラの焦点距離の逆数であり、 $\alpha_1 \sim \alpha_{N+1}$ は顔の特徴点と瞳中心点の各奥行きである。式 (8-2) において、下付き文字の t は現時点 (現フレーム) を示し、下付き文字の $t-1$ は一時刻前 (前フレーム) を示し、 $A^{(1)}$ は現状態と一時刻前の状態との顔の運動特性を表す状態遷移行列であり、 $\mu_t^{(1)}$ は現状態と一時刻前の状態とのノイズを表す。特に、 $\mu_t^{(1)}$ は、ゼロの平均と分散共分散行列で構成されるガウシアン分布のランダムベクトルである。

【0038】

また、推定部 11 では、眼球の運動モデルの状態変数である q ベクトルを式 (9-1) で定義する。 q ベクトルは、2 自由度の回転パラメータで構成される。さらに、推定部 11 では、眼球の運動モデルの状態空間表現を式 (9-2) で定義する。

【0039】

【数 9】

$$\bar{q} = \begin{pmatrix} \phi \\ \theta \end{pmatrix} \cdots (9-1)$$

$$\bar{q}_t = A^{(2)} \bar{q}_{t-1} + \mu_t^{(2)} \cdots (9-2)$$

\bar{q} : 眼球の運動モデルの状態変数

\bar{q}_t : 現時点での眼球の運動モデルの状態変数

\bar{q}_{t-1} : 一時刻前の眼球の運動モデルの状態変数

$A^{(2)}$: 状態遷移行列

$\mu_t^{(2)}$: ノイズ (ゼロの平均と $Q^{(2)}$ で構成されるガウシアン分布のランダムベクトル)

$Q^{(2)}$: ノイズの分散共分散行列

式 (9-1) において、 (θ, ϕ) は眼球の回転を表す 2 つの回転角である (図 4 参照)。式 (9-2) において、 $A^{(2)}$ は現状態と一時刻前の状態との眼球の運動特性を表す状態遷移行列であり、 $\mu_t^{(2)}$ は現状態と一時刻前の状態とのノイズを表す。

【0040】

式 (1) と式 (6) から判るように、顔の特徴点と瞳の中心点の 3 次元構造は初期フレームの画像座標を求めて定義する面では同じであるが、瞳の中心点の 3 次元構造にはそれに動的な画像上の移動量 (θ, ϕ) が加えられている。そこで、この眼球座標系 (X_e, Y_e, Z_e) が動くのに対応するために、図 5 に示すように、顔の運動モデルと眼球の運動モデルとを結合し、一点に集中させる DCCGM を生成する。図 5 に示す O_t ベクトル ($t = 1, 2, \dots$) は、顔の運動モデルの状態変数 p ベクトルと眼球の運動モデルの状態変数 q ベクトルを結合した観測ベクトルである。 O_t ベクトルは、顔の特徴点 ($i = 1, 2, \dots, N$) 及び瞳の中心点 ($i = N+1, N+2$) の全てのモデル点を画像面 (CCD) 上に投影した座標で構成される。

【0041】

推定部 11 では、式 (8-1)、式 (8-2) で示す顔の運動モデルと式 (9-1)、式 (9-2) で示す眼球の運動モデルから DCCGM の状態空間表現を式 (10-1) で定義する。さらに、推定部 11 では、観測ベクトルである O_t ベクトルを式 (10-2) で定義する。 O_t ベクトルは、 $(N+2)$ 個分の o_i ベクトルを含んでいる。

【0042】

【数 1 0】

$$\bar{X}_t = \begin{pmatrix} \bar{p}_t \\ \bar{q}_t \end{pmatrix} = \begin{pmatrix} A^{(1)} & 0 \\ 0 & A^{(2)} \end{pmatrix} \bar{X}_{t-1} + \begin{pmatrix} \mu_t^{(1)} \\ \mu_t^{(2)} \end{pmatrix} \quad \dots (10-1)$$

$$\bar{O}_t = \begin{pmatrix} \bar{o}_1 \\ \bar{o}_2 \\ \vdots \\ \bar{o}_{N+1} \\ \bar{o}_{N+2} \end{pmatrix} = f(\bar{X}_t) + v_t \quad \dots (10-2)$$

\bar{X}_t : 現時点での Dynamic Converging-Connected Gaze Model の状態変数

\bar{X}_{t-1} : 一時刻前の Dynamic Converging-Connected Gaze Model の状態変数

\bar{O}_t : 全ての特徴点の Dynamic Converging-Connected Gaze Model
に対する画像面 (CCD) 上に投影した観測ベクトル

\bar{o}_i : Dynamic Converging-Connected Gaze Model の各モデル点を
画像面 (CCD) 上に投影した観測ベクトル

$f()$: 観測関数

v_t : ノイズ

式 (10-1) において、 X_t ベクトルは、顔の運動モデルと眼球の運動モデルの状態変数を示す p_t ベクトルと q_t ベクトルからなる DCCGM の状態変数である。式 (10-2) において、 $f()$ は観測関数であり、 v_t はノイズを表す。

【0 0 4 3】

推定部 11 では、式 (1) で定義した顔の特徴点 ($i = 1, 2, \dots, N$) と式 (6) で定義した瞳の中心点 ($i = N+1, N+2$) の剛性変換 (回転と並進移動) である式 (11) を顔が動いた場合の 3 次元構造として定義する。

【0 0 4 4】

【数 1 1】

$$\begin{pmatrix} x_{ci} \\ y_{ci} \\ z_{ci} \\ 1 \end{pmatrix} = T_c^h \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} = \begin{pmatrix} R_c & t_c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} \quad \dots (11)$$

$i = 1, 2, \dots, N, N+1, N+2$

x_{ci}, y_{ci}, z_{ci} : 顔が動いた場合の顔の特徴点及び瞳の中心点を
カメラ座標上に投影した座標

T_c^h : 顔座標系のカメラ座標系に対する変換 (4×4 の行列)

R_c : 顔座標系のカメラ座標系に対する回転 (3×3 の回転行列)

t_c : 顔座標系のカメラ座標系に対する並進移動 (3×1 の並進行列)

式 (11) において、 T_c^h はカメラ座標系 (X_c, Y_c, Z_c) から顔座標系 (X_h, Y_h, Z_h) への変換であり、 R_c はカメラ座標系に対する顔の回転を示す 3×3 の回転行列であり、 t_c はカメラ座標系に対する顔の位置移動を示す 3×1 の並進行列である (図 6 参照)。この顔の動きには、制限がなく、回転行列 R_c と並進行列 t_c によってあらゆる動きを表すことができる。また、式 (11) では、顔の特徴点 ($i = 1, 2, \dots, N$) のときには式 (1) で示す (x_i, y_i, z_i) を用い、瞳の中心点 ($i = N+1, N+2$)

のときには式 (6) で示す (x_i, y_i, z_i) を用いる。

【0045】

さらに、推定部 11 では、式 (11) で示す顔の特徴点及び瞳の中心点 (x_{ci}, y_{ci}, z_{ci}) を画像面 (CCD) 上に投影した o_i ベクトル (観測ベクトル) を式 (12) で定義する。

【0046】

【数 12】

$$\bar{o}_i = \begin{pmatrix} \frac{x_{ci}}{1 + \beta z_{ci}} \\ \frac{y_{ci}}{1 + \beta z_{ci}} \end{pmatrix} \dots (12)$$

$i = 1, 2, \dots, N, N+1, N+2$

$\bar{o}_i: x_{ci}, y_{ci}, z_{ci}$ を画像面 (CCD) 上に投影した観測ベクトル

このように、推定部 11 では、式 (10-2) で示す観測ベクトル O_t から状態変数 X_t を推定するために、顔が動いた場合の 3 次元構造と観測ベクトル o_i を定義している。

【0047】

そして、推定部 11 では、式 (10-2) の観測ベクトル O_t に対して拡張カルマンフィルタリングを施し、式 (10-1) の状態変数である X_t ベクトル (すなわち、顔の運動モデルの状態変数である p ベクトル及び眼球の運動モデルである q ベクトル) の各パラメータ ($\alpha_1 \sim \alpha_{N+2}$ 、 β 、 θ 、 ϕ 等) を推定する。この観測過程では、式 (12) の観測ベクトル o_i 及び式 (11) の三次元構造を用いる。

【0048】

さらに、推定部 11 では、フィルタリングによって推定した各パラメータからカメラ座標系 (X_c, Y_c, Z_c) に対する視線座標系 (X_g, Y_g, Z_g) への変換 T_c^g を求める。変換 T_c^g は、式 (13) に示すように、カメラ座標系 (X_c, Y_c, Z_c) から顔座標系 (X_h, Y_h, Z_h) への変換 T_c^h と顔座標系 (X_h, Y_h, Z_h) から視線座標系 (X_g, Y_g, Z_g) への変換 T_h^g との積である。変換 T_c^h は、式 (11) に示すように、顔の動きに応じて回転行列 R_c 及び並進行列 t_c が設定される。変換 T_h^g は、式 (2) に示すように、 p ベクトル及び q ベクトルの各パラメータの推定値を用いて設定される。

【0049】

【数 13】

$$T_c^g = T_c^h T_h^g = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \dots (13)$$

$$R = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \dots (13-1)$$

T_c^g : 視線座標系のカメラ座標系に対する変換 (4×4 の行列)

T_c^h : 顔座標系のカメラ座標系に対する変換 (4×4 の行列)

T_h^g : 視線座標系の顔座標系に対する変換 (4×4 の行列)

R : 視線座標系のカメラ座標系に対する回転 (3×3 の回転行列)

t : 視線座標系のカメラ座標系に対する並進移動 (3×1 の並進行列)

式 (13) において、 R はカメラ座標系 (X_c, Y_c, Z_c) に対する瞳の中心点の回転を示す 3×3 の回転行列であり、 t はカメラ座標系 (X_c, Y_c, Z_c) に対する瞳の中心点の位置移動を示す 3×1 の並進行列である (図 6 参照)。

【0050】

そして、推定部 11 では、並進行列 t で示される位置を始点として回転行列 R の 3 番目の列ベクトル $(R_{13}, R_{23}, R_{33})^T$ で示す方向を視線方向とする。この視線方向は、左

右の眼で同じ方向である。ちなみに、回転行列 R の 3 番目の列ベクトルは、視線座標系 (X_g, Y_g, Z_g) における Z 軸方向である。

【0051】

このように、画像処理装置 3 では、顔の運動モデルと眼球の運動モデルとを結合した DCCGM を生成し、DCCGM に対する拡張カルマンフィルタによって静的パラメータ及び動的パラメータを推定する。そして、画像処理装置 3 では、その各パラメータの推定値からカメラ座標系 (X_c, Y_c, Z_c) から視線座標系 (X_g, Y_g, Z_g) への変換行列を求め、視線方向を求める。

【0052】

図 7 に、画像処理システム 1 によって実際に推定した視線方向を顔画像に描いた一例を示す。この例では、 320×240 の画像サイズを有する動画像を用いている。図 7 では、黒矢印で顔姿勢の方向を示し、白矢印で視線方向を示し、大きい白丸で瞳の外周を示し、多数の小さい白丸で顔の特徴点を示している。図 7 (a) の例では、顔姿勢が左方向であり、視線が顔姿勢より中心よりの左方向である。図 7 (b) の例では、顔姿勢が右方向であり、視線が左方向である。

【0053】

画像処理システム 1 (画像処理装置 3) によれば、実際に起こりうる瞳の動きを数学式でモデリングするとともに静的パラメータや動的パラメータを直接フィルタリングしているので、視線方向を高精度に推定することができる。また、画像処理システム 1 では、単眼カメラで制限のない顔の動きに対して視線の推定が可能である。さらに、画像処理システム 1 では、最も一般的な透視投影変換をカメラモデルに使用しているので、応用範囲が広い。また、画像処理システム 1 では、非接触式で視線の推定が可能であり、高価なカメラシステムや装置を必要としない。

【0054】

以上、本発明に係る実施の形態について説明したが、本発明は上記実施の形態に限定されることなく様々な形態で実施される。

【0055】

例えば、本実施の形態ではコンピュータでアプリケーションプログラムを実行することによって画像処理装置を構成したが、視線を推定するための専用の画像処理装置で構成してもよいし、あるいは、人の視線情報を必要とする装置 (例えば、ロボット、自動車の安全装置) に組み込まれる構成でもよい。

【0056】

また、本実施の形態では状態変数を推定するために拡張カルマンフィルタを用いたが、他のフィルタやフィルタ以外の手段等によって推定してもよい。

【図面の簡単な説明】

【0057】

【図 1】 本実施の形態に係る画像処理システムの構成図である。

【図 2】 カメラ座標系である。

【図 3】 顔が固定の場合のカメラ座標系、顔座標系、眼球座標系、視線座標系である。

。

【図 4】 眼球モデルである。

【図 5】 Dynamic Converging-Connected Gaze Model である。

【図 6】 顔が動いた場合のカメラ座標系、顔座標系、眼球座標系、視線座標系である。

。

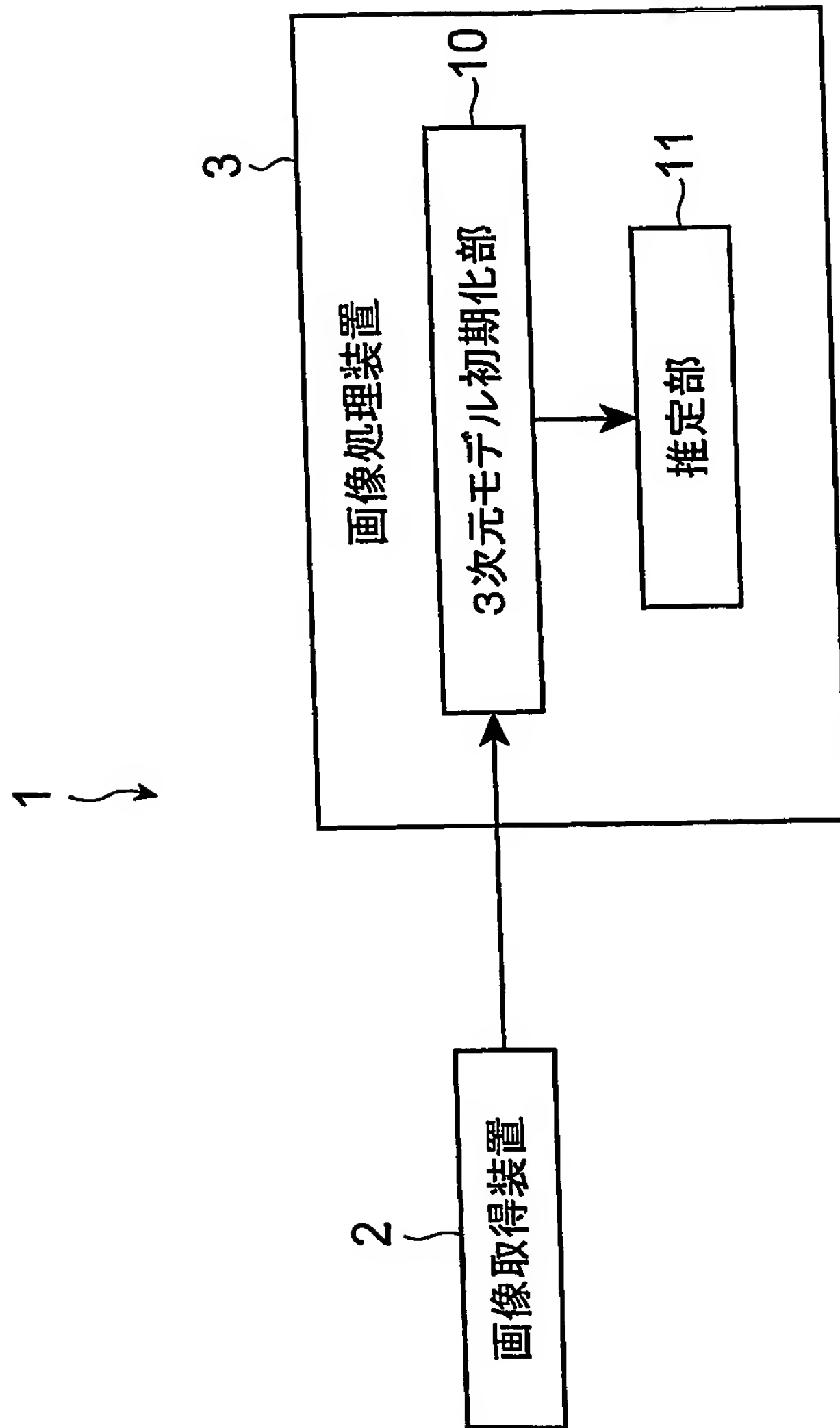
【図 7】 図 1 の画像処理装置において視線を推定した場合の顔の画像例であり、(a) が顔姿勢が左方向で視線が左方向の場合であり、(b) が顔姿勢が右方向で視線が左方向の場合である。

【符号の説明】

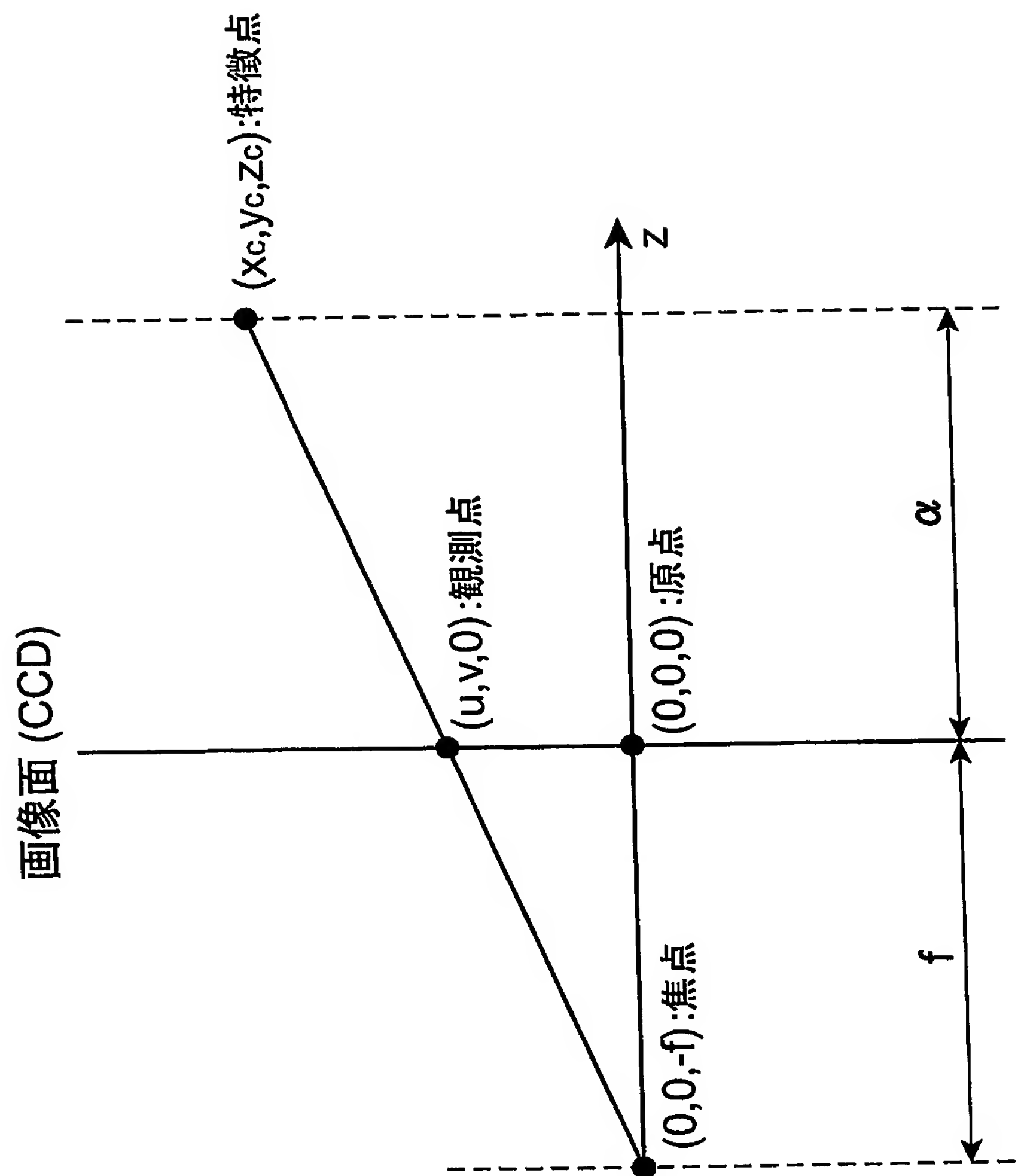
【0058】

1...画像処理システム、2...画像取得装置、3...画像処理装置、10...3次元モデル初期化部、11...推定部

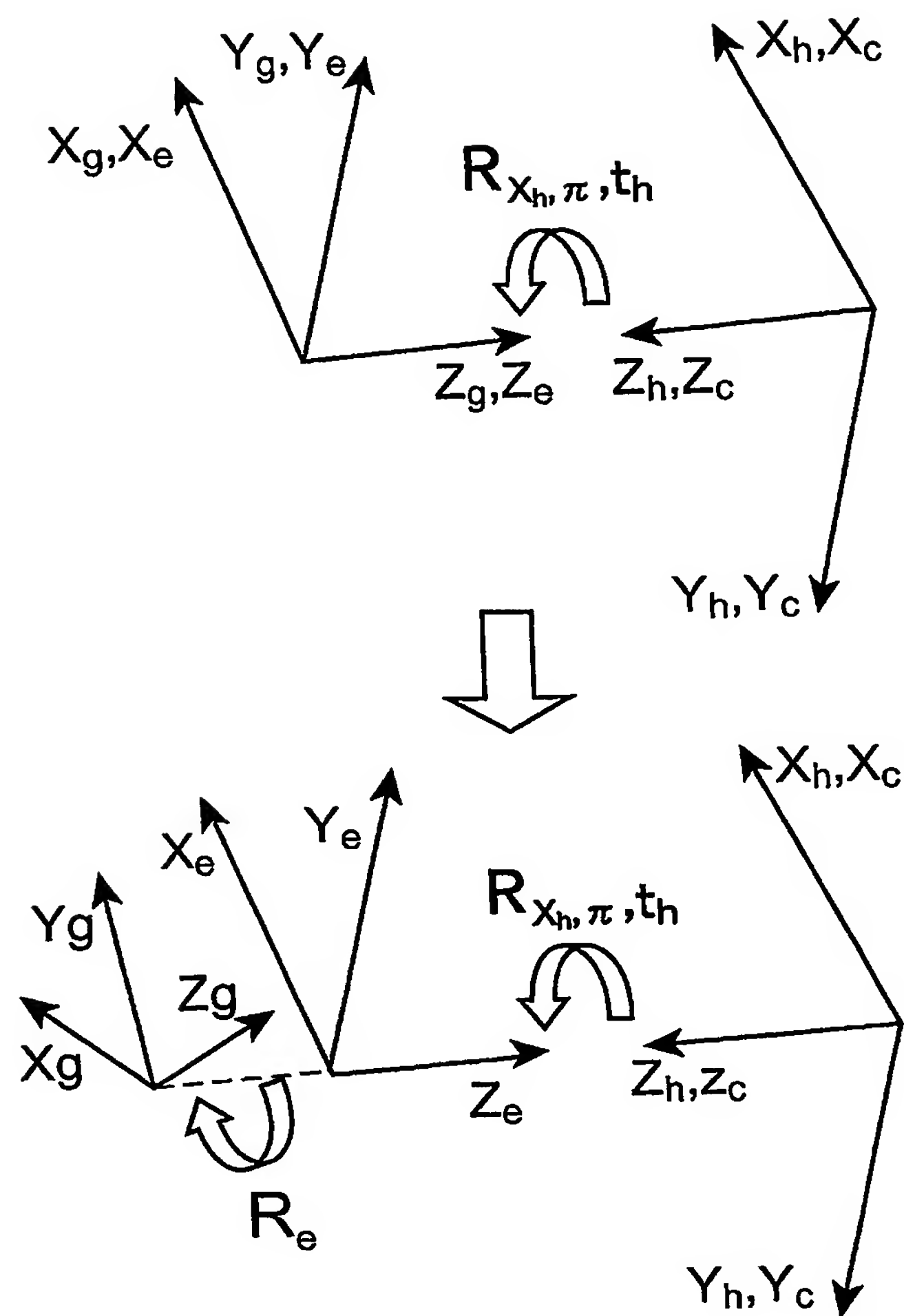
【書類名】 図面
【図 1】



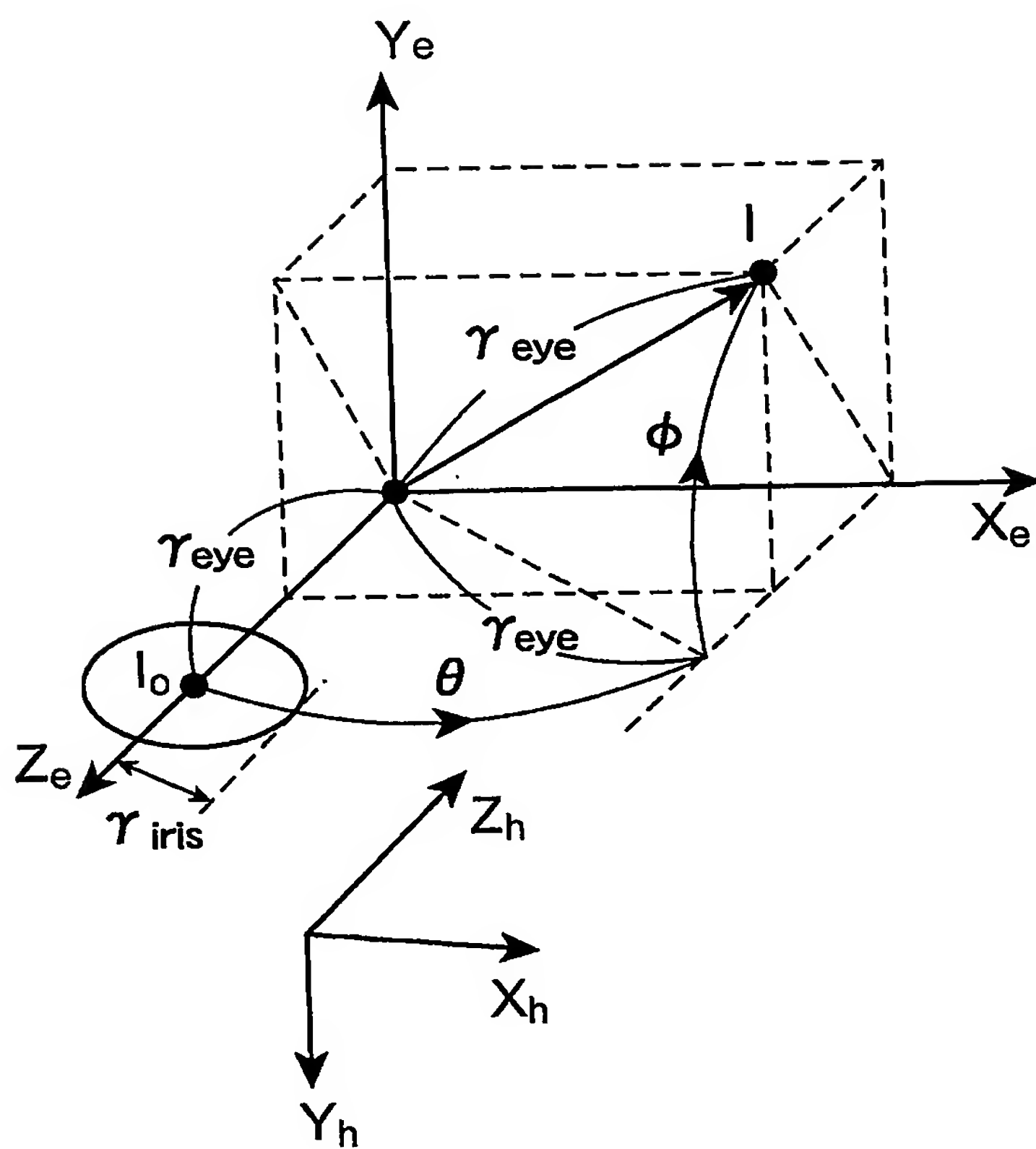
【図 2】



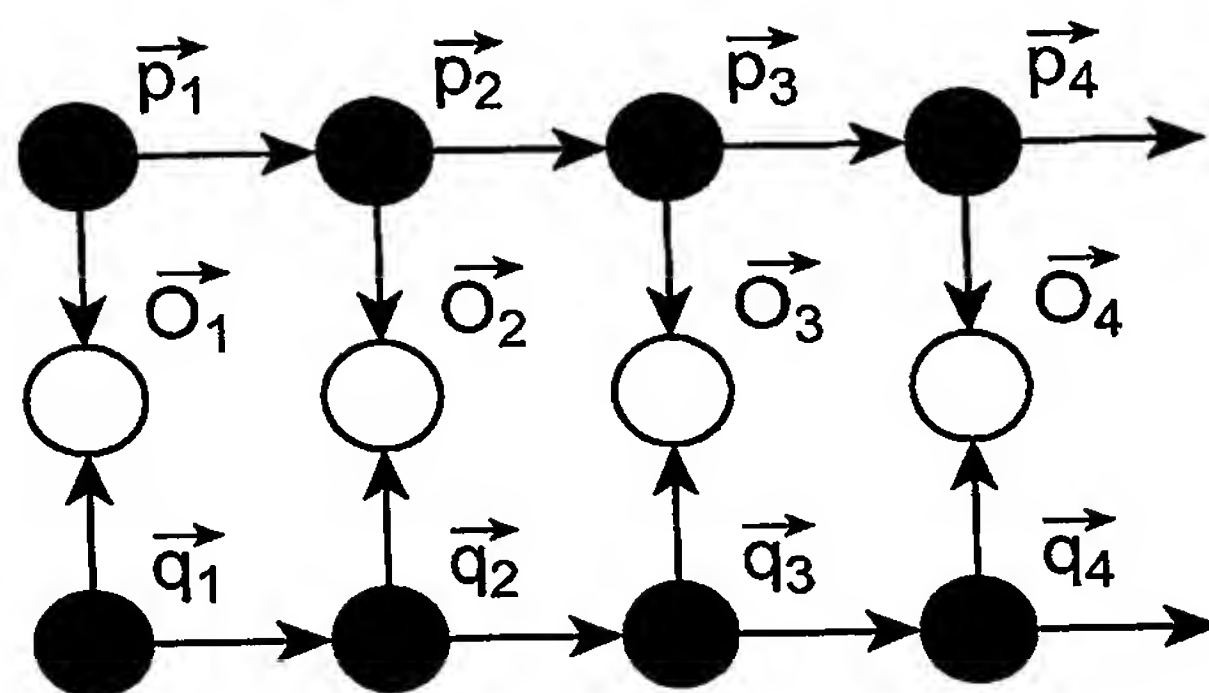
【図 3】



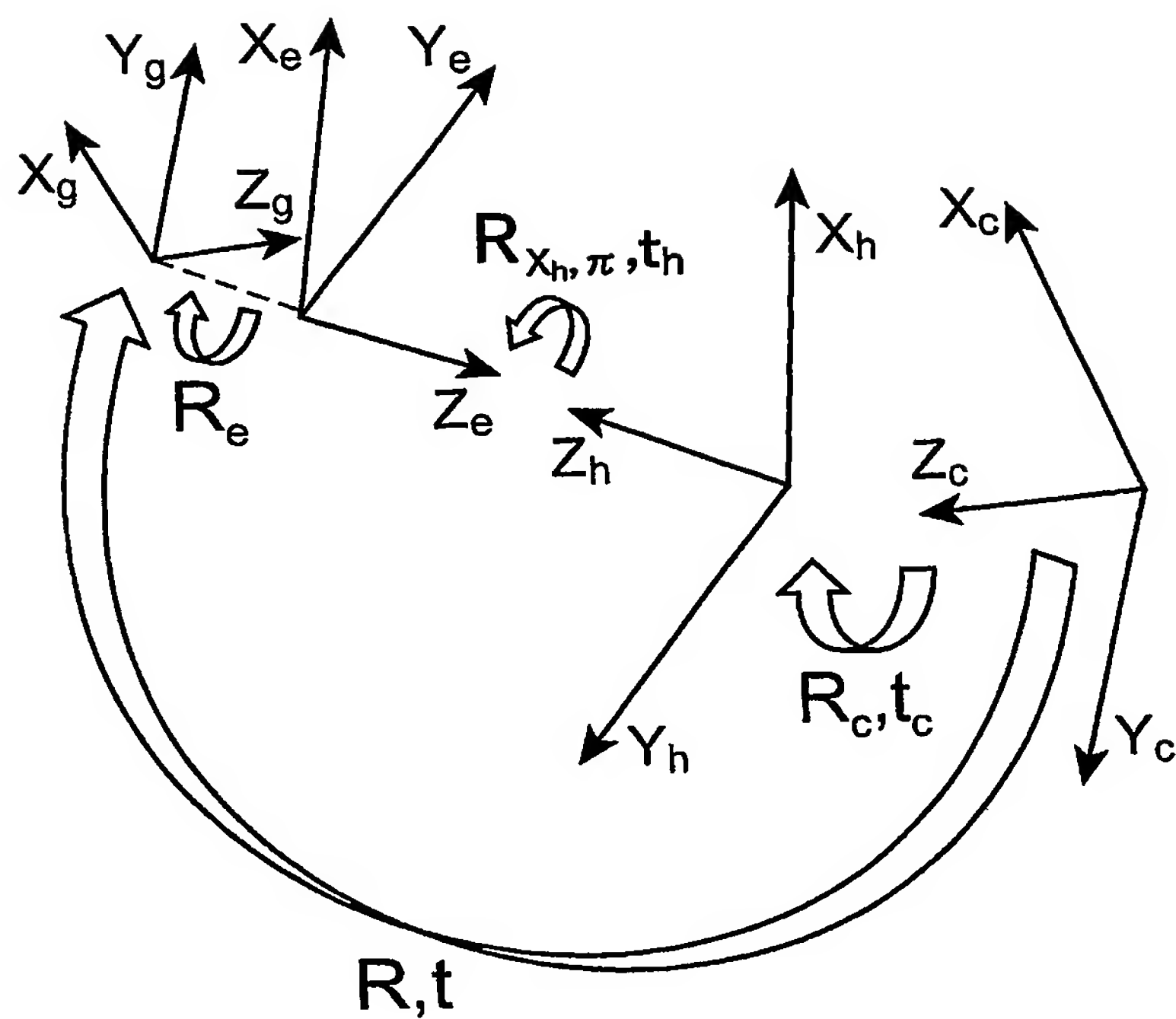
【図 4】



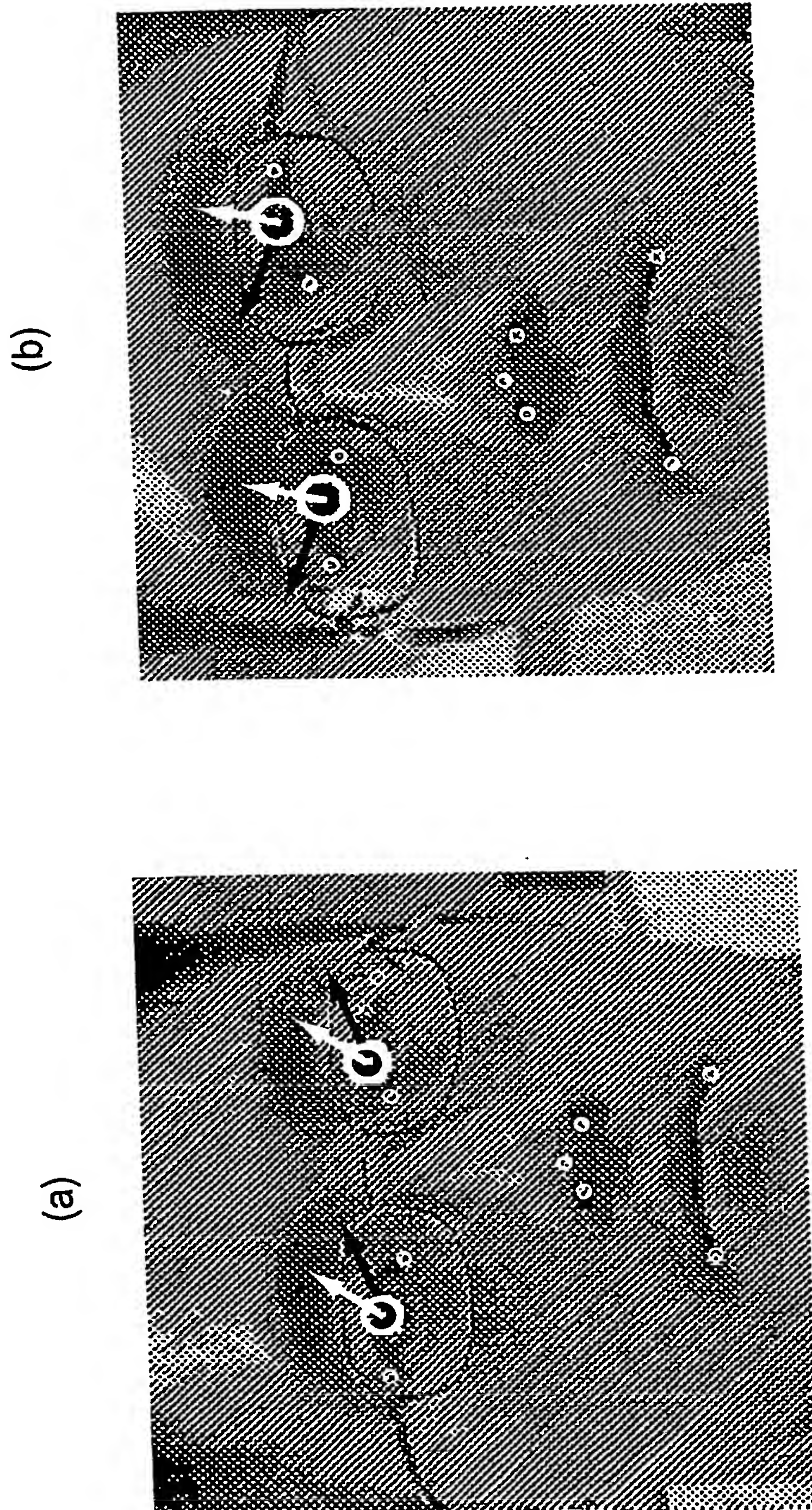
【図 5】



【図 6】



【図 7】



【書類名】 要約書

【要約】

【課題】 視線を高精度に求めることができる画像処理装置を提供することを課題とする

。【解決手段】 単眼カメラで顔を撮像し、撮像した動画像から視線を求める画像処理装置であって、顔画像上の瞳中心の 3 次元構造を静的パラメータと動的パラメータで定義し、静的パラメータ及び動的パラメータを推定することによって視線を求めることを特徴とする。

【選択図】 図 1

特願 2 0 0 3 - 2 7 3 5 9 0

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 3 2 0 7]

1. 変更年月日

1 9 9 0 年 8 月 2 7 日

[変更理由]

新規登録

住 所

愛知県豊田市トヨタ町1番地

氏 名

トヨタ自動車株式会社

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☒ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.